

Comparative analysis of genomes of 12 species of *Drosophila*

Tatiana Astakhova¹, Dmitry Malko², Vsevolod Makeev², Mikhail Roytberg¹

¹*Institute of Mathematical Problems of Biology, Russian Academy of Sciences, Puschino, Russia,*
mroytberg@impb.psn.ru

²*Vavilov Institute of General Genetics RAN 119991, Moscow, Gubkina, 3, Russia*

We have performed comparative analysis of neighborhoods of exon borders in complete genomes of genus *Drosophila*. For each exon border we have picked up two-side neighborhood, length of each flank equals 38 bp. Thus we have considered 8 types of fragments (coding and non-coding flanks of conservative dinucleotide of acceptors sites, the same for donor sites, start- and stop-codons). The dataset was prepared using programs [1] and Pro-Frame [2].

Following characteristics were calculated for each type of fragments t , position i and genome g : frequency $PN[t, i, x, g]$ of each nucleotide x ; frequency $PD[t, i, xy, g]$; of each dinucleotides xy ; Likelihood ratio $LR[t, i, xy, g] = PD[t, i, xy, g] / (PN[t, i, x, g] \cdot PN[t, i, y, g])$ of each dinucleotides xy ; for each type of fragments we have calculated also mean values of the characteristics (notation: $MPN[t, x, g]$, $MPD[t, xy, g]$, $MLR[t, xy, g]$,.. All data are available at <http://lpm.org.ru/~mroytberg/DrosophilaSITE.ZIP>.

The most significant results are as follows. (1) Frequencies of A and T nucleotides in exonic flanks in *D. willstoni* genome is greater than in other genomes, see Tab. 1, 2. It is known [3] that the preference of codons in *D. willstoni* differs significantly from other types of *Drosophila*, it is consistent with our data. (2) Likelihood ratio of dinucleotide GC in the genomes of *D. mojavensis*, *D. virilis*, *D. grimshawi* is greater than in other genomes; (3) Likelihood ratio for conventional conservative dinucleotide AG in the intron neighborhood acceptor sites is lower than 1 for all genomes, this is in correspondence with the known data for other species. (4) Likelihood ratio for dinucleotide TA in all exonic neighborhoods is lower than 1 for all genomes.

1. W.J..Kent. BLAT -. (2002) Genome Research, vol. 12: 656-664.

2. A.A.Mironov, P.S.Novichkov, M.S.Gelfand. (2001) Bioinformatics, 17:13-15.

3 Saverio Vicario, Etsuko N Moriyama and Jeffrey R. Powell. Codon usage in twelve species of *Drosophila*, (2007) *BMC Evolutionary Biology*, 7:226

a

b

	A	C	G	T	SPECIES	A	C	G	T
DMEL	26.26	22.86	11.18	39.71	DMEL	24.64	27.19	26.14	22.03
DSIM	26.10	23.16	11.10	39.64	DSIM	24.39	27.51	26.22	21.89
DSEC	26.09	23.24	11.15	39.52	DSEC	24.18	27.70	26.30	21.83
DYAK	26.00	23.42	11.05	39.53	DYAK	24.25	27.65	26.50	21.60
DERE	25.68	24.30	11.82	38.20	DERE	24.24	27.97	26.48	21.32
DANA	26.52	21.55	9.76	42.17	DANA	24.77	27.40	25.77	22.07
DPSE	21.53	26.04	12.38	40.05	DPSE	24.18	27.18	26.64	21.99
DPER	22.11	25.13	12.26	40.51	DPER	24.21	27.07	26.38	22.33
DWIL	23.44	19.00	9.57	48.00	DWIL	26.49	23.37	24.08	26.05
DMOJ	24.47	21.05	10.45	44.03	DMOJ	25.27	25.05	25.44	24.23
DVIR	25.25	19.50	11.75	43.49	DVIR	25.35	25.40	25.73	23.53
DGRI	25.68	18.38	11.31	44.62	DGRI	25.13	24.50	25.93	24.44

Tab.1. Average nucleotide composition of intronic (a) and exonic (b) neighborhoods of acceptor sites for different species of *Drosophila*.

a

b

SPECIES	A	C	G	T	SPECIES	A	C	G	T
DMEL	26.52	26.59	25.49	21.40	DMEL	31.30	19.68	18.54	30.48
DSIM	26.04	26.88	25.64	21.45	DSIM	31.13	20.05	18.82	30.00
DSEC	26.16	26.95	25.67	21.22	DSEC	30.91	20.04	18.93	30.11
DYAK	26.05	27.18	25.57	21.19	DYAK	30.65	20.40	18.39	30.56
DERE	25.82	27.37	25.88	20.93	DERE	29.91	21.11	19.27	29.72
DANA	26.05	26.83	25.18	21.94	DANA	30.16	19.72	17.90	32.22
DPSE	26.42	26.56	25.52	21.50	DPSE	27.59	24.90	20.76	26.76
DPER	26.68	26.07	25.28	21.97	DPER	28.58	24.14	20.23	27.05
DWIL	29.42	21.65	23.15	25.78	DWIL	35.17	18.02	15.91	30.89
DMOJ	27.51	25.14	24.51	22.85	DMOJ	33.10	19.30	17.35	30.24
DVIR	27.24	25.23	24.87	22.65	DVIR	33.06	20.86	16.67	29.41
DGRI	27.67	25.49	24.03	22.81	DGRI	34.28	19.29	15.03	31.40

Tab.2. Average nucleotide composition of exonic (a) and intronic (b) neighborhoods of donor sites for different species of *Drosophila*.